

# Gaussian information matrix for Wiener model identification\*

Kaushik Mahata<sup>†1</sup> and Johan Schoukens<sup>‡2</sup>

<sup>1</sup>Department of Electrical Engineering, University of Newcastle, Callaghan, NSW-2308, Australia.

<sup>2</sup>Department ELEC, Vrije Universiteit, Brussel, Building K, Pleinlaan 2, 1050, Brussels, Belgium.

October 13, 2015

## Abstract

We present a closed form expression for the information matrix associated with the Wiener model identification problem under the assumption that the input signal is a stationary Gaussian process. This expression holds under quite generic assumptions. We allow the linear sub-system to have a rational transfer function of arbitrary order, and the static nonlinearity to be a polynomial of arbitrary degree. We also present a simple expression for the determinant of the information matrix. The expressions presented herein has been used for optimal experiment design for Wiener model identification.

**Keywords:** Wiener model identification, Information matrix, Gaussian input, Determinant.

---

\*The research is supported in part by the Australian Research Council under the contract DP130103909, by the Fund for Scientific Research (FWO-Vlaanderen), by the Flemish Government (Methusalem), the Belgian Government through the Inter university Poles of Attraction (IAP VII) Program, and by the ERC advanced grant SNLSID, under contract 320378.

<sup>†</sup>Email: Kaushik.Mahata@newcastle.edu.au

<sup>‡</sup>Email: Johan.Schoukens@vub.ac.be

# 1 Introduction

In this paper we present an expression of the information matrix associated with the Wiener model identification problem under a generic setting. Our analysis allows a rational model the linear subsystem, and a polynomial nonlinearity. The rational transfer function can be of arbitrary order and a nonlinearity with arbitrary polynomial order can be allowed. The analysis assumes Gaussian stationary input process.

This work is mainly motivated by the experiment design problem. The standard approach in experiment design is to choose the input excitation in order to optimize some monotonic function of the information matrix [1–3, 8, 9, 11]. When the linear subsystem  $G$  of the underlying Wiener system has a generic rational transfer function, it has an infinite impulse response. Consequently, the information matrix  $\mathbf{J}$  becomes a function of the higher order joint moments of the input process  $u(t), u(t-1), u(t-2), \dots$ . It is very challenging to optimize any criterion of  $\mathbf{J}$  with respect to the all higher order moments of potentially infinitely many consecutive samples of the input  $u(t)$ . In fact, it is quite difficult to just compute  $\mathbf{J}$ . Firstly, the available formulae for calculating higher order moments are quite challenging to program. More importantly, the complexity of the resulting algorithm typically grows exponentially with the length of the impulse response of  $G$  [12]. In fact, to the best of our knowledge no previous authors have considered handling this issue when  $G$  is not a finite impulse response system. Even when a finite impulse response system is considered in the literature, the order of the system have been restricted to 4 or less. In fact, when compared with the experiment design literature for linear system identification [1–3, 8, 9, 11], the line of research in the nonlinear input design has undergone a significant paradigm shift. Most of the preliminary studies reported so far [6, 7, 10, 13, 16], have considered a deterministic setting. Among these the multi-level excitation approach [4–6, 13] appears to be popular lately. These deterministic methods do have their limitations. The multi-level approach is often not tractable when we increase the number of levels. The majority of these methods are unable to handle IIR-type non-linear systems.

We show when the input process is Gaussian there is a simple algorithm to compute  $\mathbf{J}$ . This analysis also reveals some interesting mathematical structures, that allows us to parameterize the set of all admissible information matrices with a finite number of parameters. Hence the experiment designer needs to solve only a finite dimensional problem. See [14] for the details of how the expressions presented herein can be used in experiment design.

## 2 Information matrix and its determinant

In this section we present our main findings about the information matrix  $\mathbf{J}$  and its determinant. We start in Section 2.1 with the basic notation and introduce the formal problem setting. In particular, we introduce a generalized framework for setting up the constraint to ensure unique identifiability of the Wiener model. Next in Section 2.2 we list the main results. In particular we use a state space representations of underlying transfer functions. We believe this approach simplifies the analysis, and illuminates the underlying mathematical structures to a significant extent.

## 2.1 Model parameterization and identifiability

A Wiener system is a cascade of a linear time invariant system followed by a static nonlinearity. We assume that the linear sub-system has a rational transfer function

$$G(z, \boldsymbol{\theta}) = \frac{g_0 + g_1 z^{-1} + \cdots + g_n z^{-n}}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}}, \quad (1)$$

parameterized by the parameter vector  $\boldsymbol{\theta}$  defined as

$$\boldsymbol{\theta} = [a_1 \ \cdots \ a_n \ g_1 \ \cdots \ g_n \ g_0]^\top. \quad (2)$$

The output of the linear model is denoted by  $w$ :

$$w(t, \boldsymbol{\theta}) = G(z, \boldsymbol{\theta})u(t). \quad (3)$$

The static nonlinearity is modeled by a polynomial  $\wp$  of order  $m$ :

$$\wp(x, \bar{\boldsymbol{\alpha}}) = \alpha_0 + \alpha_1 x + \cdots + \alpha_m x^m,$$

parameterized by the vector of polynomial coefficients

$$\bar{\boldsymbol{\alpha}} = [\alpha_0 \ \alpha_1 \ \cdots \ \alpha_m]^\top.$$

Therefore, the Wiener model equation takes the form

$$M(\boldsymbol{\vartheta}, \mathbf{u}_t) = \wp\{G(z, \boldsymbol{\theta})u(t), \bar{\boldsymbol{\alpha}}\}. \quad (4)$$

It is tempting to choose  $\boldsymbol{\vartheta} = [\bar{\boldsymbol{\alpha}}^\top \ \boldsymbol{\theta}^\top]^\top$ . But this parameterization fails to ensure unique identifiability. We cannot allow all the components of  $\boldsymbol{\theta}$  and  $\bar{\boldsymbol{\alpha}}$  to vary freely while remaining independent of each other. The reason is straightforward. The transfer operator between  $u$  and  $y$  does not change by dividing  $G$  by a scalar  $\varrho \neq 0$ , and multiplying  $\alpha_k$  by  $\varrho^k$  for all  $k = 1, 2, \dots, m$ . For this reason we must impose some additional constraint on the parameters. In this paper we allow varying the static gain of  $G$  freely, and impose a normalization constraint on  $\bar{\boldsymbol{\alpha}}$ .

**Assumption 1** *There is a known vector*

$$\mathbf{v} = [v_0 \ v_1 \ \cdots \ v_m]^\top \quad (5)$$

*such that*

$$\alpha_0 v_0 + \alpha_1 v_1 + \cdots + \alpha_m v_m = 1, \quad (6)$$

*where  $v_\ell \neq 0$  for some known  $\ell \in \{1, 2, \dots, m\}$ .*

The choice of  $\ell$  is often governed by the prior knowledge on the type of nonlinearity. Typically  $\ell \neq 0$ , because it is often the case that  $\alpha_0 = 0$ . For an odd (even) nonlinearity  $\ell$  must be an odd (even) number. In our experience, the choice of  $\ell$  does not influence the asymptotic large sample accuracy of the estimated model.

**Example 1** *It is common to take  $\mathbf{v} = (0, 1, \dots, 0)$  or  $\mathbf{v} = (0, \dots, 0, 1)$ . Another possibility would be to take  $\mathbf{v} = (1, \dots, 1)$  implying  $\wp(1) = 1$ . Note that the choice  $\mathbf{v} = (1, 0, \dots, 0)$  is forbidden. It leads to a model that is not identifiable.*

Since  $v_\ell \neq 0$  under Assumption 1, we can rewrite (6) as

$$\alpha_\ell = \frac{1}{v_\ell} \left\{ 1 - \sum_{\substack{k=0 \\ k \neq \ell}}^m v_k \alpha_k \right\}. \quad (7)$$

Equation (7) can be built into the identification algorithm. We do not identify  $\alpha_\ell$  separately, but express it using (7). We define the parameter vector

$$\boldsymbol{\alpha} := [\alpha_{i_1} \ \cdots \ \alpha_{i_m}]^\top, \quad (8)$$

where the indices  $i_k \in \{0, 1, \dots, m\}$  are chosen such that  $i_k \neq \ell$  for all  $k$ , and  $i_k \neq i_j$  whenever  $k \neq j$ . Note that mapping  $k \rightarrow i_k$  is quite flexible, and we need not impose any further restriction on this mapping. The identification algorithm estimates

$$\boldsymbol{\vartheta} = [\boldsymbol{\alpha}^\top \ \boldsymbol{\theta}^\top]^\top$$

from the data. Defining

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -v_{i_1}/v_\ell \\ 0 & 1 & \cdots & 0 & -v_{i_2}/v_\ell \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -v_{i_m}/v_\ell \end{bmatrix},$$

$$\mathbf{P} = [\mathbf{e}_{i_1} \ \cdots \ \mathbf{e}_{i_m} \ \mathbf{e}_\ell]^\top, \quad (9)$$

with  $\mathbf{e}_k$  denoting the  $k+1$  th column of  $(m+1) \times (m+1)$  identity matrix, it can be verified from (7) that

$$[\boldsymbol{\alpha}^\top \ \alpha_\ell]^\top = \mathbf{P} \bar{\boldsymbol{\alpha}} = \mathbf{L}^\top \boldsymbol{\alpha}. \quad (10)$$

## 2.2 Main theoretical results

Let  $\mathbf{a} = [a_1 \ \cdots \ a_n]^\top$ , and  $\mathbf{g} = [g_1 \ \cdots \ g_n]^\top$ . Then we can write (1) as

$$G(z, \boldsymbol{\theta}) = g_0 + (\mathbf{g} - \mathbf{a}g_0)^\top (z\mathbf{I} - \mathbf{A}_1)^{-1} \mathbf{b}_1, \quad (11)$$

where  $(\mathbf{A}_1, \mathbf{b}_1)$  is in controllable canonical form, *i.e.*

$$\mathbf{A}_1 = \begin{bmatrix} -a_1 & \cdots & -a_{n-1} & -a_n \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (12)$$

Note that we can impose the structure (11) and (12) without any loss of generality. We make the following assumption throughout the paper, where  $\boldsymbol{\theta}$  denotes the true value of  $\boldsymbol{\theta}$ .

**Assumption 2**  $G(z, \mathring{\theta})$  is asymptotically stable. Consequently, all the eigenvalues of  $\mathring{\mathbf{A}}_1$  (which denotes the true value of  $\mathbf{A}_1$ ) are located inside the unit disc in the complex plane. In addition, the state space realization (11) is minimal.

**Lemma 1** Define the matrices  $\mathbf{A}, \mathbf{b}, \mathbf{c}$  and  $\bar{\mathbf{C}}$  as

$$\begin{aligned} \bar{\mathbf{C}} &= \begin{bmatrix} \mathbf{I} & -\mathring{g}_0 \mathbf{I} & 0 \\ 0 & \mathbf{I} & 0 \\ 0 & -\mathring{\mathbf{a}}^\top & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{0}_{n \times 1} \\ \mathbf{0}_{n \times 1} \\ 1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} \mathbf{0}_{n \times 1} \\ \mathring{\mathbf{g}} \\ \mathring{g}_0 \end{bmatrix}, \\ \mathbf{A} &= \bar{\mathbf{C}} \begin{bmatrix} \mathring{\mathbf{A}}_1 & -\mathbf{b}_1(\mathring{\mathbf{g}} - \mathring{\mathbf{a}}\mathring{g}_0)^\top & \mathbf{0}_{n \times 1} \\ \mathbf{0}_{n \times n} & \mathring{\mathbf{A}}_1 & \mathbf{b}_1 \\ \mathbf{0}_{1 \times n} & \mathbf{0}_{1 \times n} & 0 \end{bmatrix} \bar{\mathbf{C}}^{-1}, \end{aligned} \quad (13)$$

where  $\mathring{\mathbf{A}}_1, \mathring{g}_0$ , etc are obtained by setting  $\boldsymbol{\theta} = \mathring{\boldsymbol{\theta}}$  in  $\mathbf{A}_1, g_0$ , etc. Consider the stochastic process  $\mathbf{x}$  which is given in state space form as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{b}u(t). \quad (14)$$

Then  $w(t, \boldsymbol{\theta}) = \mathbf{c}^\top \mathbf{x}(t)$ , and

$$\mathbf{v}_t = \begin{bmatrix} \mathbf{L}\mathbf{P}\mathbf{z}(t, \mathring{\boldsymbol{\theta}}) \\ \mathbf{x}(t)\boldsymbol{\alpha}_2^\top \mathbf{z}(t, \mathring{\boldsymbol{\theta}}) \end{bmatrix}, \quad (15)$$

where we define

$$\mathbf{z}(t, \boldsymbol{\theta}) := [1 \quad w(t, \boldsymbol{\theta}) \quad \{w(t, \boldsymbol{\theta})\}^2 \quad \cdots \quad \{w(t, \boldsymbol{\theta})\}^m]^\top, \quad (16)$$

$$\boldsymbol{\alpha}_2 = [\mathring{\alpha}_1 \quad 2\mathring{\alpha}_2 \quad \cdots \quad m\mathring{\alpha}_m \quad 0]^\top, \quad (17)$$

with  $\mathring{\alpha}_k$  being the true value of  $\alpha_k$ .

**Proof:** See Section A. ■

**Remark 1** Lemma 1 does cover the case when  $G$  is of finite impulse response type, i.e.,

$$G(z, \boldsymbol{\theta}) = g_0 + g_1 z^{-1} + \cdots + g_n z^{-n}.$$

In this case  $\boldsymbol{\theta} = [\mathbf{g}^\top \quad g_0]^\top$ , and  $\mathbf{a} = 0$ . The expressions (11) and (12) still hold with  $\mathbf{a} = 0$ . While finding a realization of  $G_1$  we do not need to consider the derivatives with respect to  $\mathbf{a}$ . As a result we get

$$\mathbf{A} = \begin{bmatrix} \mathring{\mathbf{A}}_1 & \mathbf{b}_1 \\ \mathbf{0}_{1 \times n} & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{0}_{n \times 1} \\ 1 \end{bmatrix},$$

$\bar{\mathbf{C}} = \mathbf{I}$  and  $\mathbf{c} = \boldsymbol{\theta}$ .

The consequence of Lemma 1 is that  $\mathbf{J} = \mathbf{E}\{\mathbf{v}_t \mathbf{v}_t^\top\}$  is a function of the moments of the state vector  $\mathbf{x}$ . For the purpose of setting up an input design problem we can parameterize  $\mathbf{J}$  in terms of the moments of the random vector  $\mathbf{x}$ . In particular, when  $u(t)$  is Gaussian, then so is  $\mathbf{x}(t)$ . Hence for a Gaussian input  $\mathbf{J}$  is a function of the mean and the covariance matrix of  $\mathbf{x}(t)$ . As the next Theorem reveals, we can obtain a closed form expression for  $\mathbf{J}$ .

**Assumption 3** *The input process  $u(t)$  is stationary Gaussian with mean  $\eta_u$ .*

Under Assumption 3,  $\mathbf{x}$  is a Gaussian random vector with mean

$$\boldsymbol{\eta} := \mathbb{E}\{\mathbf{x}(t)\} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{b} \eta_u. \quad (18)$$

Let us define

$$\boldsymbol{\Sigma} = \mathbb{E}\{[\mathbf{x}(t) - \boldsymbol{\eta}][\mathbf{x}(t) - \boldsymbol{\eta}]^\top\}. \quad (19)$$

Consequently,  $\mathbf{c}^\top \mathbf{x}(t)$  is a Gaussian random variable such that

$$\gamma := \mathbb{E}\{\mathbf{c}^\top \mathbf{x}(t)\} = \mathbf{c}^\top \boldsymbol{\eta}. \quad (20a)$$

$$\sigma := \mathbb{E}\{\mathbf{c}^\top \mathbf{x}(t) - \gamma\}^2 = \mathbf{c}^\top \boldsymbol{\Sigma} \mathbf{c}. \quad (20b)$$

In the rest of the paper we denote

$$\boldsymbol{\Lambda} := \mathbb{E}\{\mathbf{z}(t, \hat{\boldsymbol{\theta}})[\mathbf{z}(t, \hat{\boldsymbol{\theta}})]^\top\}.$$

**Remark 2** It is possible to express  $\boldsymbol{\Sigma}$  as well in terms of  $\mathbf{A}$ ,  $\mathbf{b}$ , and the power spectral density  $\Phi$  of  $u$ . However, we postpone that for a while. We first express  $\mathbf{J}$  in terms of  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\eta}$ , and later connect  $\Phi$  with  $\boldsymbol{\Sigma}$ . This approach suits the purpose of input design, where it is simpler to work with a parameterization of  $\boldsymbol{\Sigma}$  than to work with  $\Phi$  directly.

**Remark 3** The correlation matrix  $\boldsymbol{\Lambda}$  can be given entirely as a function of the mean  $\gamma$  and variance  $\sigma$  of  $\mathbf{c}^\top \mathbf{x}(t)$ . Many different ways are used in the literature to express the moments of the scalar valued normal density. There are some explicit formulae for smaller orders. We find it convenient to use a recursive formula in the implementation. Let us denote  $\mu_k(\gamma, \sigma) := \mathbb{E}\{(\mathbf{c}^\top \mathbf{x})^k\}$ . So  $\mu_k$  is a function of  $\sigma$  and  $\mu$ . Then  $\mu_k(\gamma, \sigma)$  satisfies the recursion [15, Chapter 5]:

$$\mu_k(\gamma, \sigma) = \gamma^k + \frac{k(k-1)}{2} \int_0^\sigma \mu_{k-2}(\tau, \sigma) d\tau. \quad (21)$$

Note that the recursion (21) needs to be carried out separately for even and odd values of  $k$ . For even valued  $k$  one can initialize the recursion with  $\mu_0(\gamma, \sigma) = 1$ , and for the odd values of  $k$  we initialize with  $\mu_1(\gamma, \sigma) = \gamma$ . This allows us to form

$$\boldsymbol{\Lambda} = \begin{bmatrix} \mu_0(\gamma, \sigma) & \mu_1(\gamma, \sigma) & \cdots & \mu_m(\gamma, \sigma) \\ \mu_1(\gamma, \sigma) & \mu_2(\gamma, \sigma) & \cdots & \mu_{m+1}(\gamma, \sigma) \\ \vdots & \vdots & & \vdots \\ \mu_m(\gamma, \sigma) & \mu_{m+1}(\gamma, \sigma) & \cdots & \mu_{2m+1}(\gamma, \sigma) \end{bmatrix}.$$

Since  $\mathbf{x}$  is Gaussian, all the moments of  $\mathbf{x}$  can be expressed as functions of  $\boldsymbol{\eta}$  and  $\boldsymbol{\Sigma}$ . This allows us to derive manageable expressions of  $\mathbf{J}$  as a function of  $\boldsymbol{\eta}$  and  $\boldsymbol{\Sigma}$ . This is shown next.

**Theorem 1** *Define*

$$\boldsymbol{\alpha}_1 = [0 \quad \hat{\alpha}_1 \quad 2\hat{\alpha}_2 \quad \cdots \quad m\hat{\alpha}_m]^\top, \quad \boldsymbol{\beta} = \boldsymbol{\alpha}_2^\top \boldsymbol{\Lambda} \boldsymbol{\alpha}_2, \quad (22)$$

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sigma} & -\frac{\gamma}{\sigma} \\ 0 & 1 \end{bmatrix}, \quad \mathbf{F} := [\mathbf{\Sigma} \mathbf{c} \quad \boldsymbol{\eta}], \quad \mathbf{H} = \begin{bmatrix} \beta\sigma & 0 \\ 0 & 0 \end{bmatrix}.$$

Partition  $\mathbf{J}$  as

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{11} & \mathbf{J}_{21}^\top \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix},$$

where  $\mathbf{J}_{11}$  is of size  $m \times m$ , while  $\mathbf{J}_{22}$  is of size  $(2n+1) \times (2n+1)$ . Then

$$\begin{aligned} \mathbf{J}_{11} &= \mathbf{L}_1 \mathbf{\Lambda} \mathbf{L}_1^\top, \\ \mathbf{J}_{21} &= \mathbf{F} \mathbf{Q} \mathbf{L}_2 \mathbf{\Lambda} \mathbf{L}_1^\top \\ \mathbf{J}_{22} &= \mathbf{F} \mathbf{Q} (\mathbf{L}_2 \mathbf{\Lambda} \mathbf{L}_2^\top - \mathbf{H}) \mathbf{Q}^\top \mathbf{F}^\top + \beta \mathbf{\Sigma}, \end{aligned} \tag{23}$$

where

$$\begin{aligned} \mathbf{L}_1 &:= \mathbf{L} \mathbf{P}, \\ \mathbf{L}_2 &:= \begin{bmatrix} \boldsymbol{\alpha}_1^\top \\ \boldsymbol{\alpha}_2^\top \end{bmatrix} = \begin{bmatrix} 0 & \dot{\alpha}_1 & 2\dot{\alpha}_2 & \cdots & m\dot{\alpha}_m \\ \dot{\alpha}_1 & 2\dot{\alpha}_2 & \cdots & m\dot{\alpha}_m & 0 \end{bmatrix}. \end{aligned} \tag{24}$$

**Proof:** See Appendix B. ■

**Remark 4** Expressions given by Theorem 1 allow us to compute  $\mathbf{J}$  in a simple way. To the best of our knowledge there is no similar expressions in the literature allowing this computational advantage.

The matrices  $\mathbf{Q}, \mathbf{H}, \mathbf{\Lambda}, \mathbf{L}_1, \mathbf{L}_2$  and  $\beta$  share an interesting property. They depend only on the true parameter vector  $\dot{\boldsymbol{\vartheta}}$  and the second order statistics (consisting of  $\gamma$  and  $\sigma$ ) of the stochastic process  $w(t, \dot{\boldsymbol{\vartheta}}) = \mathbf{c}^\top \mathbf{x}(t)$ . These quantities remain constant so long  $\gamma$  and  $\sigma$  remain constant, even though the input power spectral density (and thus  $\mathbf{\Sigma}$ ) may vary. This is due to the fact that the estimation accuracy of the static nonlinearity depends only on the amplitude distribution of  $w(t)$ , regardless of  $\mathbf{\Sigma}$  (or equivalently,  $\Phi$ ). This observation plays a key role in the sequel, and is formalized via the following definition.

**Definition 1** A quantity is called  $w$ -dependent if it is a function of  $\dot{\boldsymbol{\vartheta}}$ ,  $\sigma$  and  $\gamma$  only.

The expressions given in Theorem 1 may not seem appealing from the point of view of setting up an optimization problem for input design that can be solved in a tractable manner. The next result is more attractive in that regard.

**Theorem 2** The determinant of  $\mathbf{J}$  is given by

$$\det(\mathbf{J}) = \frac{\beta^{2n} r_1^2}{\sigma} \det(\mathbf{J}_{11}) \det(\mathbf{\Sigma}). \tag{25}$$

where  $r_1 = \boldsymbol{\alpha}_1^\top \mathbf{v} (\mathbf{v}^\top \mathbf{\Lambda}^{-1} \mathbf{v})^{-1/2}$ .

**Proof:** See Appendix C ■

**Remark 5** The expression of  $\det(\mathbf{J})$  in (25) has some nice structure. The factor

$$f := \beta^{2n} r_1^2 \det(\mathbf{J}_{11}) / \sigma \quad (26)$$

is  $w$ -dependent, and remains constant when the statistics of  $w(t, \boldsymbol{\theta}_0)$  remain invariant. On the other hand it is well-known from the literature on the input design for linear systems that we can parameterize  $\det(\boldsymbol{\Sigma})$  in a convex manner using a finite number of parameters. When the mean  $\eta_u$  of the input is kept fixed, then the above facts let us solve the D-optimal design problem for Wiener models via an one dimensional search in  $\sigma$ . To emphasize the  $w$ -dependence of  $f$  we write it as  $f(\gamma, \sigma)$ . When we consider a situation where  $\gamma$  is fixed and known, then we write it as  $f(\sigma)$ .

**Remark 6** Note that  $\mathbf{J}$  is singular when  $r_1 = 0$ . This means that the normalization of the form described in Assumption 1 ensures identifiability (and thus a non-singular information matrix) only when

$$0 \neq \boldsymbol{\alpha}_1^\top \mathbf{v} = v_1 \dot{\alpha}_1 + 2v_2 \dot{\alpha}_2 + \cdots + mv_m \dot{\alpha}_m, \quad (27)$$

see the definition of  $r_1$  in the statement of Theorem 2. We can easily construct a case where (27) fails to hold. That is  $\mathbf{v} = (1, 0, \dots, 0)$ . It is straightforward to see why this choice leads to lack of identifiability: it still allows us to simultaneously vary the gain of the linear subsystem and the factors  $\{\alpha_k\}_{k=1}^m$ , while the constraint (6) is satisfied.

By imposing the constraint (6) we restrict the search space to the hyperplane

$$\mathcal{H} = \left\{ (\alpha_0, \alpha_1, \dots, \alpha_m) : \sum_{k=0}^m \alpha_k v_k = 1 \right\}$$

By assumption,  $(\dot{\alpha}_0, \dot{\alpha}_1, \dots, \dot{\alpha}_m) \in \mathcal{H}$ . The model is identifiable when  $\mathcal{H}$  intersects with the manifold

$$\mathcal{M} = \{(\dot{\alpha}_0, \varrho \dot{\alpha}_1, \dots, \varrho^m \dot{\alpha}_m) : \varrho \neq 0\}$$

only at the point  $(\dot{\alpha}_0, \dot{\alpha}_1, \dots, \dot{\alpha}_m)$ , which corresponds to  $\varrho = 1$ . We have local identifiability at  $(\dot{\alpha}_0, \dot{\alpha}_1, \dots, \dot{\alpha}_m)$  only if  $\mathcal{M}$  is not oriented along  $\mathcal{H}$  at  $(\dot{\alpha}_0, \dot{\alpha}_1, \dots, \dot{\alpha}_m)$ , i.e.,  $\varrho = 1$ . In other words, we do not want the directional derivative  $(0, 2\dot{\alpha}_1, \dots, m\dot{\alpha}_m) =: \boldsymbol{\alpha}_1$  of  $\mathcal{M}$  at  $\varrho = 1$  to be perpendicular to  $\mathbf{v}$ , which is identical to (27).

### 3 Conclusions

We have presented several new results on the analysis of Wiener model identification using Gaussian input processes. One of the main results in this paper is Theorem 1, which gives a closed form expression of the associated information matrix  $\mathbf{J}$ . This expression holds under very generic assumptions on the model structure. In addition, unlike other similar formulae available in the literature, our expression for  $\mathbf{J}$  is easy to compute. This aspect makes it attractive in input design problems. Theorem 2 gives a simple expression for the determinant of  $\mathbf{J}$ . These expressions are particularly useful in experiment design, see [14] for details.



## A Proof of Lemma 1

By definition of  $\mathbf{P}$  in (9) we have  $\mathbf{P}\mathbf{P}^\top = \mathbf{I}$ . Using this in (10) gives

$$\bar{\alpha} = \mathbf{P}^\top \mathbf{L}^\top \alpha. \quad (28)$$

Using (28) and the definition of  $\mathbf{z}(t, \boldsymbol{\theta})$  in (16) in (4) we have

$$M(\boldsymbol{\vartheta}, \mathbf{u}_t) = \bar{\alpha}^\top \mathbf{z}(t, \boldsymbol{\theta}) = \alpha^\top \mathbf{L} \mathbf{P} \mathbf{z}(t, \boldsymbol{\theta}). \quad (29)$$

Hence

$$\frac{\partial M(\boldsymbol{\vartheta}, \mathbf{u}_t)}{\partial \alpha} = \mathbf{L} \mathbf{P} \mathbf{z}(t, \boldsymbol{\theta}). \quad (30)$$

Also using the definition of  $\mathbf{z}(t, \boldsymbol{\theta})$  in (16) and differentiating  $M(t, \boldsymbol{\vartheta})$  in (29) with respect to  $\boldsymbol{\theta}$  we get

$$\begin{aligned} \frac{\partial M(\boldsymbol{\vartheta}, \mathbf{u}_t)}{\partial \boldsymbol{\theta}} &= \frac{\partial w(t, \boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}} \alpha^\top \mathbf{L} \mathbf{P} \begin{bmatrix} 0 \\ 1 \\ 2w(t, \boldsymbol{\theta}) \\ \vdots \\ m\{w(t, \boldsymbol{\theta})\}^{m-1} \end{bmatrix} \\ &= \frac{\partial w(t, \boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}} \alpha_2^\top \mathbf{z}(t, \boldsymbol{\theta}), \end{aligned} \quad (31)$$

where the last equality follows from the definition of  $\alpha_2$  in (17) and the definition of  $\mathbf{z}(t, \boldsymbol{\theta})$  in (16). The proof for the expression of  $\mathbf{v}_t$  in (15) will be complete if we can show

$$\frac{\partial w(t, \boldsymbol{\vartheta})}{\partial \boldsymbol{\theta}} = \frac{\partial G(z, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} u(t) = \mathbf{x}(t). \quad (32)$$

This is done next. By differentiating  $G$  with respect to  $\mathbf{a}$ ,  $\mathbf{g}$  and  $g_0$  we get

$$\begin{aligned} \frac{\partial G(z, \boldsymbol{\theta})}{\partial \mathbf{a}} &= -(z\mathbf{I} - \mathbf{A}_1)^{-1} \mathbf{b}_1 g_0 \\ &\quad - (z\mathbf{I} - \mathbf{A}_1)^{-1} \mathbf{b}_1 (\mathbf{g} - \mathbf{a} g_0)^\top (z\mathbf{I} - \mathbf{A}_1)^{-1} \mathbf{b}_1, \end{aligned} \quad (33a)$$

$$\frac{\partial G(z, \boldsymbol{\theta})}{\partial \mathbf{g}} = (z\mathbf{I} - \mathbf{A}_1)^{-1} \mathbf{b}_1, \quad (33b)$$

$$\frac{\partial G(z, \boldsymbol{\theta})}{\partial g_0} = 1 - \mathbf{a}^\top (z\mathbf{I} - \mathbf{A}_1)^{-1} \mathbf{b}_1. \quad (33c)$$

Using (33) and (13) it can be verified by direct calculations that

$$\frac{\partial G(z, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = (\mathbf{I} - \mathbf{A} z^{-1})^{-1} \mathbf{b}, \quad (34)$$

implying (32).

To show  $w(t, \boldsymbol{\theta}) = \mathbf{c}^\top \mathbf{x}(t)$  verify from (11) and (33) that

$$G(z, \boldsymbol{\theta}) = [\mathbf{0}^\top \quad \mathbf{g}^\top \quad g_0] \frac{\partial G(z, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{c}^\top (\mathbf{I} - \mathbf{A} z^{-1})^{-1} \mathbf{b}.$$

## B Proof of Theorem 1

Since  $\Sigma$  is positive definite,  $\Sigma \mathbf{c} \neq 0$ . Hence there exists a full column rank  $(2n+1) \times (2n)$  matrix  $\mathbf{C}$  such that the column space of  $\mathbf{C}$  is the orthogonal complement of  $\Sigma \mathbf{c}$ , i.e.,  $\mathbf{C}^\top \Sigma \mathbf{c} = 0$ . Hence

$$\begin{bmatrix} \mathbf{c}^\top \\ \mathbf{C}^\top \end{bmatrix} \Sigma \begin{bmatrix} \mathbf{c} & \mathbf{C} \end{bmatrix} = \begin{bmatrix} \sigma & 0 \\ 0 & \Sigma_1 \end{bmatrix}, \quad (35)$$

The block diagonal structure of the matrix in the right hand side of (35) ensures that by premultiplying  $\mathbf{x}$  by  $\begin{bmatrix} \mathbf{c} & \mathbf{C} \end{bmatrix}^\top$  we get two mutually uncorrelated components  $\mathbf{c}^\top \mathbf{x}$  and

$$\mathbf{x}_1 := \mathbf{C}^\top \mathbf{x},$$

with

$$\begin{aligned} \gamma &:= \mathbb{E}\{\mathbf{x}_1\} = \mathbf{C}^\top \boldsymbol{\eta}, \\ \Sigma_1 &:= \mathbb{E}\{[\mathbf{x}_1 - \gamma][\mathbf{x}_1 - \gamma]^\top\} = \mathbf{C}^\top \Sigma \mathbf{C}. \end{aligned} \quad (36)$$

Because  $\mathbf{x}$  is a Gaussian random vector, we conclude that  $\begin{bmatrix} \mathbf{c}^\top \mathbf{x} & \mathbf{x}_1^\top \end{bmatrix}^\top$  too is a jointly Gaussian random vector. Since uncorrelated Gaussian variables are independent,  $\mathbf{c}^\top \mathbf{x}$  and  $\mathbf{x}_1$  are mutually independent.

Define the  $(m+2n+1) \times (m+2n+1)$  matrix

$$\mathbf{T} = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{c}^\top \\ 0 & \mathbf{C}^\top \end{bmatrix}, \quad (37)$$

where the identity matrix appearing in (37) in the north-west corner is of size  $m \times m$ . Premultiplying  $\mathbf{v}_t$  in (16) by  $\mathbf{T}$  we note that

$$\mathbf{T} \mathbf{v}_t = \begin{bmatrix} \mathbf{L}_1 \mathbf{z}(t, \dot{\boldsymbol{\theta}}) \\ \mathbf{c}^\top \mathbf{x}(t) \boldsymbol{\alpha}_2^\top \mathbf{z}(t, \dot{\boldsymbol{\theta}}) \\ \mathbf{C}^\top \mathbf{x}(t) \boldsymbol{\alpha}_2^\top \mathbf{z}(t, \dot{\boldsymbol{\theta}}) \end{bmatrix}. \quad (38)$$

From Lemma 1 recall that  $\mathbf{c}^\top \mathbf{x}(t) = w(t, \dot{\boldsymbol{\theta}})$ . Then from the definition of  $\mathbf{z}(t, \boldsymbol{\theta})$  in (16), the definitions  $\boldsymbol{\alpha}_1$  and  $\boldsymbol{\alpha}_2$  in (22) and (17) we have

$$\mathbf{c}^\top \mathbf{x}(t) \boldsymbol{\alpha}_2^\top \mathbf{z}(t, \dot{\boldsymbol{\theta}}) = w(t, \dot{\boldsymbol{\theta}}) \boldsymbol{\alpha}_1^\top \mathbf{z}(t, \dot{\boldsymbol{\theta}}).$$

In addition,  $\mathbf{C}^\top \mathbf{x}(t) = \mathbf{x}_1(t)$ . Hence

$$\mathbf{T} \mathbf{v}_t = \begin{bmatrix} \mathbf{L}_1 \mathbf{z}(t, \dot{\boldsymbol{\theta}}) \\ \boldsymbol{\alpha}_1^\top \mathbf{z}(t, \dot{\boldsymbol{\theta}}) \\ \mathbf{x}_1(t) \boldsymbol{\alpha}_2^\top \mathbf{z}(t, \dot{\boldsymbol{\theta}}) \end{bmatrix}. \quad (39)$$

Since  $\mathbf{x}_1(t)$  is independent of  $w(t, \dot{\boldsymbol{\theta}}) = \mathbf{c}^\top \mathbf{x}(t)$ , it is also independent of  $\mathbf{z}(t, \dot{\boldsymbol{\theta}})$ , see (16). Using this and (39) we get

$$\begin{aligned} \mathbf{T} \mathbf{J} \mathbf{T}^\top &= \mathbb{E}\{[\mathbf{T} \mathbf{v}_t][\mathbf{T} \mathbf{v}_t]^\top\} \\ &= \begin{bmatrix} \mathbf{L}_1 \boldsymbol{\Lambda} \mathbf{L}_1^\top & \mathbf{L}_1 \boldsymbol{\Lambda} \boldsymbol{\alpha}_1 & \mathbf{L}_1 \boldsymbol{\Lambda} \boldsymbol{\alpha}_2 \boldsymbol{\gamma}^\top \\ \boldsymbol{\alpha}_1^\top \boldsymbol{\Lambda} \mathbf{L}_1^\top & \boldsymbol{\alpha}_1^\top \boldsymbol{\Lambda} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_1^\top \boldsymbol{\Lambda} \boldsymbol{\alpha}_2 \boldsymbol{\gamma}^\top \\ \boldsymbol{\gamma} \boldsymbol{\alpha}_2^\top \boldsymbol{\Lambda} \mathbf{L}_1^\top & \boldsymbol{\gamma} \boldsymbol{\alpha}_2^\top \boldsymbol{\Lambda} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2^\top \boldsymbol{\Lambda} \boldsymbol{\alpha}_2 (\boldsymbol{\gamma} \boldsymbol{\gamma}^\top + \Sigma_1) \end{bmatrix}. \end{aligned} \quad (40)$$

Define the vector  $\mathbf{d}$  and the  $(2n+1) \times (2n)$  matrix  $\mathbf{D}$  by partitioning the inverse

$$\begin{bmatrix} \mathbf{c}^\top \\ \mathbf{C}^\top \end{bmatrix}^{-1} = [\mathbf{d} \ \mathbf{D}]. \quad (41)$$

Then (37) and (40) imply

$$\mathbf{J} = \begin{bmatrix} \mathbf{I} & 0 & 0 \\ 0 & \mathbf{d} & \mathbf{D} \end{bmatrix} (\mathbf{T}\mathbf{J}\mathbf{T}^\top) \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{d}^\top \\ 0 & \mathbf{D}^\top \end{bmatrix}.$$

Using expression of  $\mathbf{T}\mathbf{J}\mathbf{T}^\top$  in (40) we get

$$\mathbf{J}_{11} = \mathbf{L}_1 \mathbf{A} \mathbf{L}_1^\top, \quad (42a)$$

$$\mathbf{J}_{21} = [\mathbf{d} \ \mathbf{D}\boldsymbol{\gamma}] \mathbf{L}_2 \mathbf{A} \mathbf{L}_1^\top \quad (42b)$$

$$\mathbf{J}_{22} = [\mathbf{d} \ \mathbf{D}\boldsymbol{\gamma}] \mathbf{L}_2 \mathbf{A} \mathbf{L}_2^\top [\mathbf{d} \ \mathbf{D}\boldsymbol{\gamma}]^\top + \beta \mathbf{D} \boldsymbol{\Sigma}_1 \mathbf{D}^\top. \quad (42c)$$

Now from (35) and (41) we obtain

$$\boldsymbol{\Sigma} = [\mathbf{d} \ \mathbf{D}] \begin{bmatrix} \sigma & 0 \\ 0 & \boldsymbol{\Sigma}_1 \end{bmatrix} \begin{bmatrix} \mathbf{d}^\top \\ \mathbf{D}^\top \end{bmatrix} = \mathbf{d}\sigma\mathbf{d}^\top + \mathbf{D}\boldsymbol{\Sigma}_1\mathbf{D}^\top. \quad (43)$$

By definition of  $\mathbf{d}$  and  $\mathbf{D}$  in (41) we know

$$\begin{bmatrix} \mathbf{c}^\top \\ \mathbf{C}^\top \end{bmatrix} [\mathbf{d} \ \mathbf{D}] = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{I} \end{bmatrix},$$

and this implies  $\mathbf{C}^\top \mathbf{d} = 0$ ,  $\Rightarrow \mathbf{d} = k\boldsymbol{\Sigma}\mathbf{c}$ . In addition,  $1 = \mathbf{c}^\top \mathbf{d} = k\mathbf{c}^\top \boldsymbol{\Sigma} \mathbf{c} = k\sigma$ . Consequently,

$$\mathbf{d} = \frac{1}{\sigma} \boldsymbol{\Sigma} \mathbf{c}. \quad (44)$$

On the other hand

$$\mathbf{I} = [\mathbf{d} \ \mathbf{D}] \begin{bmatrix} \mathbf{c}^\top \\ \mathbf{C}^\top \end{bmatrix} = \mathbf{d}\mathbf{c}^\top + \mathbf{D}\mathbf{C}^\top = \frac{1}{\sigma} \boldsymbol{\Sigma} \mathbf{c} \mathbf{c}^\top + \mathbf{D}\mathbf{C}^\top, \quad (45)$$

Now multiply both sides of (45) by  $\boldsymbol{\eta}$  to get

$$\boldsymbol{\eta} - \frac{\gamma}{\sigma} \boldsymbol{\Sigma} \mathbf{c} = \mathbf{D}\boldsymbol{\gamma} \quad (46)$$

From (44) and (46) it follows that

$$[\mathbf{d} \ \mathbf{D}\boldsymbol{\gamma}] = \mathbf{F}\mathbf{Q}.$$

Now we use (43), (44), and (46) in (42) to eliminate  $\mathbf{d}$  and  $\mathbf{D}$  from the expressions of  $\mathbf{J}_{12}$  and  $\mathbf{J}_{22}$ . We have

$$\begin{aligned} \mathbf{J}_{21} &= \mathbf{F}\mathbf{Q}\mathbf{L}_2\mathbf{A}\mathbf{L}_1^\top \\ \mathbf{J}_{22} &= \mathbf{F}\mathbf{Q}\mathbf{L}_2\mathbf{A}\mathbf{L}_2^\top\mathbf{Q}^\top\mathbf{F}^\top + \beta(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{c}\mathbf{c}^\top\boldsymbol{\Sigma}/\sigma) \\ &= \beta\boldsymbol{\Sigma} + \mathbf{F}\mathbf{Q}\mathbf{L}_2\mathbf{A}\mathbf{L}_2^\top\mathbf{Q}^\top\mathbf{F}^\top - \mathbf{F}\mathbf{Q}\mathbf{H}\mathbf{Q}^\top\mathbf{F}^\top \\ &= \mathbf{F}\mathbf{Q}(\mathbf{L}_2\mathbf{A}\mathbf{L}_2^\top - \mathbf{H})\mathbf{Q}^\top\mathbf{F}^\top + \beta\boldsymbol{\Sigma}. \end{aligned}$$

## C Proof of Theorem 2

### C.1 Some Schur complement expressions

**Lemma 2** *The Schur complement  $\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top$  admits an expression*

$$\begin{aligned} \mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top \\ = \beta\boldsymbol{\Sigma} + \mathbf{F}\mathbf{Q}[\mathbf{L}_2\mathbf{v}(\mathbf{v}^\top\boldsymbol{\Lambda}^{-1}\mathbf{v})^{-1}\mathbf{v}^\top\mathbf{L}_2^\top - \mathbf{H}]\mathbf{Q}^\top\mathbf{F}^\top. \end{aligned}$$

**Proof:** In this proof we let  $\boldsymbol{\Gamma}$  be the Cholesky factor of  $\boldsymbol{\Lambda}$ , i.e.,  $\boldsymbol{\Lambda} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$ . From the expressions of  $\mathbf{J}_{11}$ ,  $\mathbf{J}_{21}$  and  $\mathbf{J}_{22}$  in Theorem 1 it follows that

$$\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top = \beta\boldsymbol{\Sigma} + \mathbf{F}\mathbf{Q}[\mathbf{L}_2\boldsymbol{\Pi}\mathbf{L}_2^\top - \mathbf{H}]\mathbf{Q}^\top\mathbf{F}^\top, \quad (47)$$

where we define

$$\begin{aligned} \boldsymbol{\Pi} &= \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\mathbf{L}_1^\top(\mathbf{L}_1\boldsymbol{\Lambda}\mathbf{L}_1^\top)^{-1}\mathbf{L}_1\boldsymbol{\Lambda} \\ &= \boldsymbol{\Gamma}[\mathbf{I} - \boldsymbol{\Gamma}^\top\mathbf{L}_1^\top(\mathbf{L}_1\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top\mathbf{L}_1^\top)^{-1}\mathbf{L}_1\boldsymbol{\Gamma}]\boldsymbol{\Gamma}^\top. \end{aligned} \quad (48)$$

However, the matrix  $\bar{\boldsymbol{\Pi}} := \mathbf{I} - \boldsymbol{\Gamma}^\top\mathbf{L}_1^\top(\mathbf{L}_1\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top\mathbf{L}_1^\top)^{-1}\mathbf{L}_1\boldsymbol{\Gamma}$  is the orthogonal projection operator onto the nullspace of  $\mathbf{L}_1\boldsymbol{\Gamma}$ .

From (5), (9) and the definition of  $\mathbf{L}_1$  in (24) verify that that  $\mathbf{L}_1\mathbf{v} = \mathbf{L}\mathbf{P}\mathbf{v} = 0$ . This means

$$\mathbf{L}_1\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{-1}\mathbf{v} = 0,$$

i.e. the vector  $\boldsymbol{\Gamma}^{-1}\mathbf{v}$  spans the one dimensional nullspace of  $\mathbf{L}_1\boldsymbol{\Gamma}$ . Hence  $\bar{\boldsymbol{\Pi}}$  is also the orthogonal projection operator onto the span of  $\boldsymbol{\Gamma}^{-1}\mathbf{v}$ . Hence

$$\bar{\boldsymbol{\Pi}} = \boldsymbol{\Gamma}^{-1}\mathbf{v}(\mathbf{v}^\top\boldsymbol{\Lambda}^{-1}\mathbf{v})^{-1}\mathbf{v}^\top\boldsymbol{\Gamma}^{-\top}.$$

Substituting this expression in (48) gives

$$\boldsymbol{\Pi} = \mathbf{v}(\mathbf{v}^\top\boldsymbol{\Lambda}^{-1}\mathbf{v})^{-1}\mathbf{v}^\top,$$

which upon substitution in (47) yields the desired result. ■

Define

$$r_i := \boldsymbol{\alpha}_i^\top\mathbf{v}(\mathbf{v}^\top\boldsymbol{\Lambda}^{-1}\mathbf{v})^{-1/2}, \quad i = 1, 2. \quad (49)$$

Note that

$$\mathbf{L}_2\mathbf{v}(\mathbf{v}^\top\boldsymbol{\Lambda}^{-1}\mathbf{v})^{-1}\mathbf{v}^\top\mathbf{L}_2^\top = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}^\top, \quad (50)$$

see the definition of  $\mathbf{L}_2$  in Theorem 1. When  $r_2 = 0$  the matrix  $\mathbf{L}_2\mathbf{v}(\mathbf{v}^\top\boldsymbol{\Lambda}^{-1}\mathbf{v})^{-1}\mathbf{v}^\top\mathbf{L}_2^\top - \mathbf{H}$  is of rank 1. Then the calculations turn out to be quite different from the case where  $r_2 \neq 0$ .

**Lemma 3** *If  $r_2 = 0$  then*

$$\det(\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top) = \frac{r_1^2}{\beta\sigma} \det(\beta\mathbf{\Sigma}), \quad (51)$$

**Proof:** When  $r_2 = 0$  then using (49), definition of  $\mathbf{Q}$  in Theorem 1 and the expressions given by Lemma 2 we get

$$\begin{aligned} & \mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top \\ &= \beta\mathbf{\Sigma} + \mathbf{F}\mathbf{Q} \begin{bmatrix} r_1^2 - \beta\sigma & 0 \\ 0 & 0 \end{bmatrix} \mathbf{Q}^\top \mathbf{F}^\top \\ &= \beta\mathbf{\Sigma} + \mathbf{F} \begin{bmatrix} \frac{1}{\sigma} & -\frac{\gamma}{\sigma} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_1^2 - \beta\sigma & 0 \\ 0 & 0 \end{bmatrix} \mathbf{Q}^\top \mathbf{F}^\top \\ &= \beta\mathbf{\Sigma} + \mathbf{F} \begin{bmatrix} r_1^2/\sigma - \beta & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma} & 0 \\ -\frac{\gamma}{\sigma} & 1 \end{bmatrix} \mathbf{F}^\top \\ &= \beta\mathbf{\Sigma} + \mathbf{F} \begin{bmatrix} r_1^2/\sigma^2 - \beta/\sigma & 0 \\ 0 & 0 \end{bmatrix} \mathbf{F}^\top \\ &= \beta\mathbf{\Sigma} + (r_1^2/\sigma^2 - \beta/\sigma)\mathbf{\Sigma}\mathbf{c}\mathbf{c}^\top\mathbf{\Sigma}. \end{aligned} \quad (52)$$

In this proof we write

$$q = r_1^2/\sigma^2 - \beta/\sigma$$

compactly. From (52) we have

$$\begin{aligned} \det(\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top) &= \det(\beta\mathbf{\Sigma} + q\mathbf{\Sigma}\mathbf{c}\mathbf{c}^\top\mathbf{\Sigma}) \\ &= \det(\beta\mathbf{\Sigma}) \det\left(\mathbf{I} + \frac{q}{\beta}\mathbf{c}\mathbf{c}^\top\mathbf{\Sigma}\right) \\ &= \det(\beta\mathbf{\Sigma}) \det\left(1 + \frac{q}{\beta}\mathbf{c}^\top\mathbf{\Sigma}\mathbf{c}\right) \\ &= \det(\beta\mathbf{\Sigma}) \det\left(1 + \frac{q\sigma}{\beta}\right) \end{aligned}$$

Substituting the value of  $q$  we get (51). ■

**Lemma 4** *Suppose that  $r_2 \neq 0$ . Then*

$$\det(\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top) = \frac{r_1^2}{\beta\sigma} \det(\beta\mathbf{\Sigma}), \quad (53)$$

$$\begin{aligned} (\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top)^{-1} &= \left[ \frac{1}{r_1^2} - \frac{1}{\beta\sigma} \left( \frac{r_2\gamma}{r_1} - 1 \right)^2 \right] \mathbf{c}\mathbf{c}^\top \\ &\quad + \left( \frac{r_2}{r_1}\mathbf{c}\boldsymbol{\eta}^\top - \mathbf{I} \right) [\beta\mathbf{\Sigma}]^{-1} \left( \frac{r_2}{r_1}\mathbf{c}\boldsymbol{\eta}^\top - \mathbf{I} \right)^\top. \end{aligned} \quad (54)$$

**Proof:** Define

$$\mathbf{B} := \left[ \mathbf{Q} \left( \frac{\mathbf{L}_2 \mathbf{v} \mathbf{v}^\top \mathbf{L}_2^\top}{(\mathbf{v}^\top \mathbf{\Lambda}^{-1} \mathbf{v})^{-1}} - \mathbf{H} \right) \mathbf{Q}^\top \right]^{-1} \quad (55)$$

Recall that  $\zeta = r_1/r_2$ . Hence from (50) we get

$$\begin{aligned} & \left( \frac{\mathbf{L}_2 \mathbf{v} \mathbf{v}^\top \mathbf{L}_2^\top}{(\mathbf{v}^\top \mathbf{\Lambda}^{-1} \mathbf{v})^{-1}} - \mathbf{H} \right)^{-1} \\ &= -\frac{1}{\beta\sigma} \begin{bmatrix} 1 & -r_1/r_2 \\ -r_1/r_2 & r_1^2/r_2^2 - \beta\sigma/r_2^2 \end{bmatrix} \\ &= -\frac{1}{\beta\sigma} \begin{bmatrix} 1 & -\zeta \\ -\zeta & \zeta^2 - \beta\sigma/r_2^2 \end{bmatrix}. \end{aligned}$$

Hence by definition of  $\mathbf{Q}$ , see Theorem 1, we get

$$\begin{aligned} \beta\mathbf{B} &= -\frac{1}{\sigma} \begin{bmatrix} \sigma & 0 \\ \gamma & 1 \end{bmatrix} \begin{bmatrix} 1 & -\zeta \\ -\zeta & \zeta^2 - \beta\sigma/r_2^2 \end{bmatrix} \mathbf{Q}^{-1} \\ &= -\frac{1}{\beta\sigma} \begin{bmatrix} \sigma & -\sigma\zeta \\ \gamma - \zeta & -\zeta(\gamma - \zeta) - \beta\sigma/r_2^2 \end{bmatrix} \begin{bmatrix} \sigma & \gamma \\ 0 & 1 \end{bmatrix} \\ &= -\frac{1}{\sigma} \begin{bmatrix} \sigma^2 & \sigma(\gamma - \zeta) \\ \sigma(\gamma - \zeta) & (\gamma - \zeta)^2 - \beta\sigma/r_2^2 \end{bmatrix} \\ &= -\begin{bmatrix} \sigma & \gamma - \zeta \\ \gamma - \zeta & \frac{(\gamma - \zeta)^2}{\sigma} - \beta/r_2^2 \end{bmatrix}. \end{aligned} \quad (56)$$

Taking determinant we have

$$\det(\beta\mathbf{B}) = -\frac{\beta\sigma}{r_2^2}. \quad (57)$$

On the other hand, recall that  $\mathbf{F} = [\mathbf{\Sigma} \mathbf{c} \quad \boldsymbol{\eta}]$ . Hence using (20) we get

$$\mathbf{F}^\top \mathbf{\Sigma}^{-1} \mathbf{F} = \begin{bmatrix} \sigma & \gamma \\ \gamma & \boldsymbol{\eta}^\top \mathbf{\Sigma}^{-1} \boldsymbol{\eta} \end{bmatrix}. \quad (58)$$

Combining (56) and (58) we get

$$\beta\mathbf{B} + \mathbf{F}^\top \mathbf{\Sigma}^{-1} \mathbf{F} = \begin{bmatrix} 0 & \zeta \\ \zeta & \boldsymbol{\eta}^\top \mathbf{\Sigma}^{-1} \boldsymbol{\eta} + \frac{\beta}{r_2^2} - \frac{(\gamma - \zeta)^2}{\sigma} \end{bmatrix} \quad (59)$$

Taking determinant we have

$$\det(\beta\mathbf{B} + \mathbf{F}^\top \mathbf{\Sigma}^{-1} \mathbf{F}) = -\zeta^2 \quad (60)$$

Now using Lemma 2 and (55) we know

$$\mathbf{J}_{22} - \mathbf{J}_{21} \mathbf{J}_{11}^{-1} \mathbf{J}_{21}^\top = \beta\mathbf{\Sigma} + \mathbf{F} \mathbf{B}^{-1} \mathbf{F}^\top \quad (61)$$

Hence

$$\begin{aligned}
& \det(\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top) \\
&= \det(\beta\boldsymbol{\Sigma} + \mathbf{F}\mathbf{B}^{-1}\mathbf{F}^\top) \\
&= \det(\beta\boldsymbol{\Sigma}) \det(\mathbf{I} + \boldsymbol{\Sigma}^{-1}\mathbf{F}(\beta\mathbf{B})^{-1}\mathbf{F}^\top) \\
&= \det(\beta\boldsymbol{\Sigma}) \det(\mathbf{I} + \mathbf{F}^\top\boldsymbol{\Sigma}^{-1}\mathbf{F}\{\beta\mathbf{B}\}^{-1}) \\
&= \frac{\det(\beta\boldsymbol{\Sigma})}{\det(\beta\mathbf{B})} \det(\beta\mathbf{B} + \mathbf{F}^\top\boldsymbol{\Sigma}^{-1}\mathbf{F}) \\
&= \det(\beta\boldsymbol{\Sigma}) \frac{r_2^2 \zeta^2}{\beta\sigma} = \det(\beta\boldsymbol{\Sigma}) \frac{r_1^2}{\beta\sigma}.
\end{aligned}$$

## C.2 Proof of the formula for $\det(\mathbf{J})$

Using Schur's determinant formula we know

$$\det(\mathbf{J}) = \det(\mathbf{J}_{11}) \det(\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top). \quad (62)$$

The result of Theorem 2 is immediate from (62) once we use the expression for  $\det(\mathbf{J}_{22} - \mathbf{J}_{21}\mathbf{J}_{11}^{-1}\mathbf{J}_{21}^\top)$  given by (53).

## References

- [1] M. Barenthin, X. Bombois, H. Hjalmarsson, and G. Scorletti. Identification for control of multivariable systems: Controller validation and experiment design via LMIs. *Automatica*, 44:3070–3078, Dec 2008.
- [2] M. Barenthin and H. Hjalmarsson. Identification and control: Joint input design and  $\mathcal{H}_\infty$  state feedback with ellipsoidal parametric uncertainty via LMIs. *Automatica*, 44(2):543–551, Feb 2008.
- [3] X. Bombois, G. Scorletti, M. Gevers, P. M. J. Van den Hof, and R. Hildebrand. Least costly identification experiment for control. *Automatica*, 42:3:1651–1662, 2006.
- [4] A. De Cock, M. Gevers, and J. Schoukens. A preliminary study on optimal input design for nonlinear systems. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 4931–4936, Dec 2013.
- [5] A. De Cock, M. Gevers, and J. Schoukens. D-optimal input design for FIR-type nonlinear systems: A dispersion based approach. 2014. Submitted for publication.
- [6] M. Forgione, X. Bombois, P. M. J. Van den Hof, and H. Hjalmarsson. Experiment design for parameter estimation in nonlinear systems based on multilevel excitation. 2014. European Control Conference.
- [7] M. Gevers, M. Caenepeel, and J. Schoukens. Experiment design for the identification of a simple Wiener system. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 7333–7338, Dec 2012.

- [8] R. Hildebrand and M. Gevers. Identification for control: optimal input design with respect to a worst-case  $\nu$ -gap cost function. *SIAM journal of control and optimization*, 41:5:1586–1608, 2003.
- [9] H. Hjalmarsson and H. Jansson. Closed loop experiment design for linear time invariant dynamical systems via LMIs. *Automatica*, 44(3):623–636, Mar 2008.
- [10] H. Hjalmarsson and J. Mårtensson. Optimal input design for identification of non-linear systems: Learning from the linear case. In *American Control Conference, 2007. ACC '07*, pages 1572–1576, July 2007.
- [11] H. Jansson and H. Hjalmarsson. Input design via LMIs admitting frequency-wise model specifications in confidence regions. *IEEE Transactions on Automatic Control*, 50(10):1534–1549, Oct 2005.
- [12] R. Kan. From moments of sum to moments of product. *Journal of multivariate analysis*, 99:542–554, 2008.
- [13] C. A. Larsson, H. Hjalmarsson, and C. R. Rojas. On optimal input design for nonlinear FIR-type systems. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 7220–7225, Dec 2010.
- [14] K. Mahata, J. Schoukens, and A. De Cock. Information matrix and D-optimal design with gaussian inputs for wiener model identification. *Automatica*, 2015. In press.
- [15] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1991.
- [16] P. E. Valenzuela, C. R. Rojas, and H. Hjalmarsson. Optimal input design for non-linear dynamic systems: A graph theory approach. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 5740–5745, Dec 2013.